(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/38767 A2

(51) International Patent Classification[7]: C12N 15/12, 5/10, C07K 14/705, 16/28, C12Q 1/68, G01N 33/577, 33/68, A01K 67/027

(21) International Application Number: PCT/US01/49696

(22) International Filing Date:
7 November 2001 (07.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/246,553 7 November 2000 (07.11.2000) US

(71) Applicant *(for all designated States except US)*: **GENETICS INSTITUTE, INC.** [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GULUKOTA, Kamalakar** [IN/US]; 3 Stout Court, Lawrenceville, NJ 08648 (US). **KAUR, Harleen** [IN/US]; 165 Mammouth Road, Lowell, MA 01854 (US). **BATES, Brian, G.** [US/US]; 16 Eldorado Road, Chelmsford, MA 01824 (US). **PAULSEN, Janet, E.** [US/US]; 43 Sherwood Road, Londonderry, NH 03053 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/38767 A2

(54) Title: NOVEL G PROTEIN-COUPLED RECEPTOR PROTEIN AND NUCLEIC ACID MOLECULES AND USES THEREFOR

(57) Abstract: Novel C5RL polypeptides, proteins, and nucleic acid molecules are disclosed. In addition to isolated, full-length C5RL proteins, the invention further provides isolated C5RL fusion proteins, antigenic peptides and anti-C5RL antibodies. The invention also provides C5RL nucleic acid molecules, recombinant expression vectors containing a nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced and non-human transgenic animals in which a C5RL gene has been introduced or disrupted. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

NOVEL G PROTEIN-COUPLED RECEPTOR PROTEIN AND NUCLEIC ACID MOLECULES AND USES THEREFOR

Related Applications

This application claims the benefit of U.S. Provisional Application Serial No. 60/246,553, filed November 7, 2000, the entire contents of which are incorporated herein by this reference.

Background of the Invention

Molecular cloning studies have shown that G protein-coupled receptors ("GPCRs") form one of the largest protein superfamilies found in nature, and it is estimated that greater than 1000 different such receptors exist in mammals. Upon binding of extracellular ligands, GPCRs interact with a specific subset of heterotrimeric G-proteins that can then, in their activated forms, inhibit or activate various effector enzymes and/or ion channels. The ligands for many of these receptors are known although there exists an ever-increasing number of GPCRs which have been identified in the sequencing of the human genome which for have no ligands have yet been identified. This latter subfamily of GPCRs is called the orphan family of GPCRs. In addition to both GPCRs with known ligands, as well as orphan GPCRs, there exist a family of GPCR-like molecules which share significant homology as well as many of the structural properties of the GPCR superfamily. For example, a family of GPCR-like proteins which arises from three alternatively-spliced forms of a gene occurring between the CD4 and triosephosphate isomerase genes at human chromosome 12p13, has been recently identified (including protein A-1, A-2, and A-3). Ansari-Lari *et al.* (1996) *Genome Res.* 6(4):314-326. Comparative sequence analysis of the syntenic region in mouse chromosome 6 has further revealed a murine homologue of at least the A-2 splice product. Ansari-Lari *et al.* (1998) *Genome Res.* 8(1):29-40.

The fundamental knowledge that GPCRs play a role in regulating that activity of virtually every cell in the human body has fostered an extensive search for modulators of such receptors for use as human therapeutics. In fact, the superfamily of GPCRs has proven to be among the most successful drug targets. Consequently, it has been recognized that the newly isolated orphan GPCRs, as well as the GPCR-like proteins, have great potential for drug discovery.

With the identification of each new GPCR, orphan GPCR, and GPCR-like protein, there exists a need for identifying the surrogate ligands for such molecules as well as for modulators of such molecules for use in regulating a variety of cellular responses.

Summary of the Invention

The present invention is based, at least in part, on the discovery of nucleic acid molecules which encode a novel family of G protein-coupled receptor-like proteins, referred to herein as the C5a receptor-like ("C5RL" family or "C5RL proteins"). The C5RL molecules of the present invention as well as C5RL ligands and/or C5RL modulators, are useful in regulating a variety of cellular processes. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding C5RL proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of C5RL-encoding nucleic acids.

In one embodiment, a C5RL nucleic acid molecule is 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical to the nucleotide sequence shown in SEQ ID NO:1 or 3, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or complement thereof. In a preferred embodiment, an isolated C5RL nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:3, or a complement thereof. In another preferred embodiment, an isolated C5RL nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:1, or a complement thereof. In yet another preferred embodiment, an isolated nucleic acid molecule has the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

In another embodiment, a C5RL nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous or identical to the amino acid sequence of SEQ ID NO:2 or an amino acid or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession No. PTA-2776. In another preferred embodiment, a C5RL nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical to the amino acid sequence of SEQ ID NO:2 or an amino acid or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession No. PTA-2776.

In another embodiment, an isolated nucleic acid molecule of the present invention encodes a C5RL protein which includes at least one, preferably two, three, four, five, six, or seven transmembrane domains. In another embodiment, an isolated nucleic acid molecule of the present invention encodes a protein which includes a 7 transmembrane receptor profile. In yet another embodiment, a C5RL nucleic acid molecule encodes a C5RL protein and is a naturally occurring nucleotide sequence.

Another embodiment of the invention features C5RL nucleic acid molecules which specifically detect C5RL nucleic acid molecules relative to nucleic acid molecules encoding non-C5RL proteins. For example, in one embodiment, a C5RL nucleic acid molecule is at least 30 nucleotides in length and hybridizes under stringent conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1 or 3, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof. Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a C5RL nucleic acid.

Another aspect of the invention provides a vector comprising a C5RL nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. The invention also provides a method for producing a C5RL protein by culturing in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a C5RL protein is produced.

Another aspect of this invention features isolated or recombinant C5RL proteins and polypeptides. In one embodiment, an isolated C5RL protein includes at least one, preferably two, three, four, five, six or seven transmembrane domains. In another embodiment, an isolated C5RL protein includes a 7 transmembrane receptor profile. In another embodiment, an isolated C5RL protein has an amino acid sequence sufficiently homologous or identical to the amino acid sequence of SEQ ID NO:2. In a preferred embodiment, a C5RL protein has an amino acid sequence at least about 70% identical to the amino acid sequence of SEQ ID NO:2. In another embodiment, a C5RL protein has the amino acid sequence of SEQ ID NO:2.

Another embodiment of the invention features an isolated C5RL protein which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 70% identical to a nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof. This invention further features an isolated C5RL protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3.

The C5RL proteins of the present invention, or biologically active portions thereof, can be operatively linked to a non-C5RL polypeptide to form C5RL fusion proteins. The invention further features antibodies that specifically bind C5RL proteins, such as monoclonal or polyclonal antibodies. In addition, the C5RL proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting C5RL expression in a biological sample by contacting the biological sample with an agent capable of detecting a C5RL nucleic acid molecule, protein or polypeptide such that the presence of a C5RL nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of C5RL activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of C5RL activity such that the presence of C5RL activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating C5RL activity comprising contacting the cell with an agent that modulates C5RL activity such that C5RL activity in the cell is modulated. In one embodiment, the agent inhibits C5RL activity. In another embodiment, the agent stimulates C5RL activity. In one embodiment, the agent is an antibody that specifically binds to a C5RL protein. In another embodiment, the agent modulates expression of C5RL by modulating transcription of a C5RL gene or translation of a C5RL mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a C5RL mRNA or a C5RL gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant C5RL protein or nucleic acid expression or activity by administering an agent which is a C5RL modulator to the subject. In one embodiment, the C5RL modulator is a C5RL protein. In another embodiment, the C5RL modulator is a C5RL nucleic acid molecule. In another embodiment, the C5RL modulator is a C5RL ligand. In yet another embodiment, the C5RL modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant C5RL protein or nucleic acid expression is an immune system disorder.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a C5RL protein; (ii) mis-regulation of said gene; and (iii) aberrant post-translational modification of a C5RL protein, wherein a wild-type form of said gene encodes an protein with a C5RL activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of a C5RL protein. In one embodiment, the invention provides a method for identifying a compound which binds to a C5RL protein which involves contacting the C5RL protein, or a cell expressing the C5RL protein with a test compound and determining whether the C5RL protein binds to the test compound. In another embodiment, the invention provides a method for identifying a compound

which modulates the activity of a C5RL protein which involves contacting a C5RL protein with a test compound, and determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the C5RL protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figures 1A-1B depicts the cDNA sequence and predicted amino acid sequence of human C5RL. The nucleotide sequence (*Figure 1A*) corresponds to nucleic acids 1 to 1374 of SEQ ID NO:1. The coding region is underlined. The amino acid sequence (*Figure 1B*) corresponds to amino acids 1 to 337 of SEQ ID NO:2. Transmembrane domains are underlined.

Figure 2 depicts a multiple sequence alignment (MSA) of the amino acid sequences of guinea pig C5R1 (SEQ ID NO:4), human C5R1 (SEQ ID NO:5), and human C5RL (SEQ ID NO:2). The 7 transmembrane receptor profiles are indicated in italics. The transmembrane domains are underlined. Asterisks indicate important conserved residues in G protein-coupled receptors. The alignment was performed using the Clustal algorithm which is part of the MEGALIGN program (*e.g.*, version 3.1.7) which is part of the DNASTAR sequence analysis software package. The pairwise alignment parameters are as follows: K-tuple = 1; Gap Penalty = 3; Window = 5; Diagonals saved = 5. The multiple alignment parameters are as follows: Gap Penalty = 10; and Gap length penalty = 10.

Detailed Description of the Invention

The present invention is based on the discovery of novel molecules, referred to herein as C5a receptor-like ("C5RL") protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence homology or identity as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin. Members of a family may also have common functional characteristics.

For example, the family of G protein-coupled receptors (GPCRs), to which the C5RL proteins of the present invention bear significant homology, comprise an N-terminal extracellular domain, seven transmembrane domains (also referred to as membrane-spanning domains), three extracellular domains (also referred to as extracellular loops), three cytoplasmic domains (also referred to as cytoplasmic loops), and a C-terminal cytoplasmic domain (also referred to as a cytoplasmic tail). Members of the GPCR family also share certain conserved amino acid residues, some of which have been determined to be critical to receptor function and/or G protein signaling. For example, GPCRs contain the following features: a conserved asparagine residue in the first transmembrane domain; a cysteine residue in the first extracellular loop which is believed to form a disulfide bond with a conserved cysteine residue in the second extracellular loop; a conserved leucine and aspartate residue in the second transmembrane domain; a conserved tryptophan and proline residue in the fourth transmembrane domain; a conserved proline residue in the fifth transmembrane; and conserved phenylalanine and proline residues in the sixth transmembrane which are commonly found as part of the motif FXXCXXP. The proline residues in the fourth, fifth and sixth transmembrane domains are thought to introduce kinks in the alpha-helices and may be important in the formation of the ligand binding pocket. The conserved residues described herein are indicated by asterisks in Figure 2.

In one embodiment, the C5RL proteins of the present invention are proteins having an amino acid sequence of about 300-370, preferably about 310-360, more preferably about 320-350, more preferably about 330-340, or about 337 amino acids in length. In another embodiment, the C5RL proteins of the present invention contain at least one transmembrane domain. As used herein, the term "transmembrane domain" includes an amino acid sequence having at least about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 amino acid residues, of which at least about 30-40%, 40-50%, 50-60% or more of the amino acid residues contain non-polar side chains, for example, alanine, valine, leucine, isoleucine, proline, phenylalanine, tryptophan, and methionine. A transmembrane domain is lipophilic in nature and spans a biological lipid membrane. For example, a transmembrane domain can be found at about amino acids 36-61 of SEQ ID NO:2. In a preferred embodiment, a C5RL protein of the present invention has more than one transmembrane domain, preferably two, three, four, five, six or seven transmembrane domains. For example, transmembrane domains can be found centered around about amino acids 50, 85, 120, 160, 215, 245 and 285 of SEQ ID NO:2 (*e.g.*, at about amino acids 71-94, 109-130, 154-172, 205-226, 234-259 and 275-297 of SEQ ID NO:2). In a particularly preferred embodiment, a C5RL protein of the present invention has seven transmembrane domains.

In another embodiment, a C5RL family member is identified based on the presence of at least one, preferably two or three cytoplasmic loops, also referred to herein as cytoplasmic domains. In another embodiment, a C5RL family member is identified based on the presence of at least one, preferably two or three extracellular loops. As defined herein, the term "loop" includes an amino acid sequence having a length of at least about 4, preferably about 5-10, more preferably about 10-15, and more preferably about 15-20, 20-25, 25-30 or more amino acid residues, and has an amino acid sequence that connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a naturally-occurring GPCR or GPCR-like protein (*e.g.*, in a C5RL protein), and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a transmembrane domain in a naturally-occurring GPCR or GPCR-like protein (*e.g.*, in a C5RL protein).

As used herein, a "cytoplasmic loop" includes an amino acid sequence located within a cell or within the cytoplasm of a cell. For example, a cytoplasmic loop is found at about amino acids 62-70, 131-153 and 227-233 of SEQ ID NO:2. Also as used herein, an "extracellular loop" includes an amino acid sequence located outside of a cell, or extracellularly. For example, an extracellular loop can be found at about amino acid residues 95-108, 173-204 and 260-274 of SEQ ID NO:2.

In another embodiment of the invention, a C5RL family member is identified based on the presence of a "C-terminal cytoplasmic domain", also referred to herein as a C-terminal cytoplasmic tail, in the sequence of the protein. As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least about 10, preferably about 10-20, and more preferably about 20-30, 30-40, or more amino acid residues and is located within a cell or within the cytoplasm of a cell. Accordingly, the N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a naturally-occurring GPCR or GPCR-like protein (*e.g.*, in a C5RL protein). For example, a C-terminal cytoplasmic domain is found at about amino acid residues 298-337 of SEQ ID NO:2.

In another embodiment, a C5RL family member is identified based on the presence of an "N-terminal extracellular domain", also referred to herein as an N-terminal extracellular loop in the amino acid sequence of the protein. As used herein, an "N-terminal extracellular domain" includes an amino acid sequence having at least about 10, preferably about 10-20, and more preferably about 20-30, 30-40 or more amino acid residues and is located outside of a cell or extracellularly. The C-terminal amino acid residue of a "N-terminal extracellular domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a naturally-occurring GPCR or GPCR-like

protein (*e.g.*, in a C5RL protein). For example, an N-terminal extracellular domain is found at about amino acid residues 1 to 35 of SEQ ID NO:2.

The intracellular, extracellular and transmembrane domains of proteins of the invention can be identified in accordance with known techniques for determination of such domains from sequence information. For example, the TopPred computer algorithm (von Heijne (1992) *J. Mol. Biol.* 225:487-494) can be used to predict the location of transmembrane domains in an amino acid sequence, domains which are described by the location of the center of the transmembrane domain, with at least 6, 7, 8, 9, 10, 11, 12, 13 or 14 transmembrane amino acids on each side of the reported central residue(s).

Accordingly in one embodiment of the invention, a C5RL family member includes at least one, preferably two, three, four, five, six or seven transmembrane domains and/or at least one, two or three cytoplasmic loops, and/or at least one, two or three extracellular loops. In another embodiment, the C5RL family member further includes an N-terminal extracellular domain and/or a C-terminal cytoplasmic domain. In another embodiment, the C5RL family member can include six transmembrane domains, three cytoplasmic loops, and two extracellular loops, or can include six transmembrane domains, three extracellular loops, and 2 cytoplasmic loops. The former embodiment can further include an N-terminal extracellular domain. The latter embodiment can further include a C-terminal cytoplasmic domain. In another embodiment, the C5RL family member can include seven transmembrane domains, three cytoplasmic loops, and three extracellular loops and can further include an N-terminal extracellular domain or a C-terminal cytoplasmic domain.

In another embodiment, a C5RL family member is identified based on the presence of at least one "7 transmembrane receptor profile", also referred to as a "Rhodopsin family sequence profile", in the protein or corresponding nucleic acid molecule. As used herein, the term "7 transmembrane receptor profile" includes an amino acid sequence having at least about 60-70, preferably about 70-80, more preferably about 80-90 amino acid residues, or at least about 90-100, 110-120, 120-130, 130-140, 140-150 or more amino acids in length and having a bit score for the alignment of the sequence to the 7tm_1 family Hidden Markov Model (HMM) of at least 20, preferably 20-30, more preferably 30-40, more preferably 40-50, 50-60, 60-70, 70-80 or greater. The 7tm_1 family HMM has been assigned the PFAM Accession PF00001 (Washington University Website, Pfam Home Page).

To identify the presence of a 7 transmembrane receptor profile in a C5RL family member, the amino acid sequence of the protein family member is searched against a database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (*e.g.*, The Sanger Center Web Server, Pfam Protein Search Software). For example, the

hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for PF00001 and score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (*e.g.*, to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3)405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.*(1990) *Methods Enzymol.* 183:146-159; Gribskov *et al.*(1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.*(1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.*(1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

For example, a search using the amino acid sequence of SEQ ID NO:2 was performed against the HMM database resulting in the identification of two 7 TM receptor profiles in the amino acid sequence of SEQ ID NO:2. The results of the search are set forth below.

```
7tm_1: domain 1 of 2, from 52 to 131: score 70.8, E = 5.3e-23

    *->GNlLVilvilrtkklrtptnifilNLAvADLLflltlppwalyylvg
       GN++V +v ++  ++r +  +++l+LAvADLL +l+lp  a+ ++ g
 52    GNAMVAWVAGKVARRR-VGATWLLHLAVADLLCCLSLPILAVPIARG 97

    gsedWpfGsalCklvtaldvvnmyaSillLtaISiD<-*
    g  +Wp+G ++C+ + ++++++myaS+llL+a+S D
 98 G--HWPYGAVGCRALPSIILLTMYASVLLLAALSAD    131


7tm_1: domain 2 of 2, from 153 to 291: score 53.9, E = 1.1e-17

    *->vvillvWvlalllslPpllfswvktveegngtlnvnvtvClidfpee
       v ++++W+lalll++P ++++ +++++++          +C +d+  +
153    VACGAAWTLALLLTVPSAIYRRLHQEHFPA------RLQCVVDYGGS 193

    stasvstwlrsyvllstlvgFllPllvilvcYtrIlrtlrkaaktll.vv
    s     + + +++ + +l+gFl Pl+  + c +  l+ + + ++ l++++
194 S-----STENAVTAIRFLFGFLGPLVAVASCHSALLCWAARRCRPLGtAI 238

    vvvFvlCWlPyfivllldtlc.lsiimsstCelervlptallvtlwLayv
    vv F++CW+Py+++ l+ t++ +++       l +   +a ++++ La +
239 VVGFFVCWAPYHLLGLVLTVAaPNS------ALLARALRAEPLIVGLALA 282

    NsclNPiIY<-*
    +sclNP++
283 HSCLNPMLF    291
```

Accordingly, in one embodiment of the invention, a C5RL protein is a human C5RL protein having at least one, preferably two 7 transmembrane receptor profiles (*e.g.*, at about amino acids 52-131 of SEQ ID NO:2 and at about amino acids 153 to 291 of SEQ ID NO:2). In another embodiment, a C5RL family members has a 7 transmembrane receptor profile or domain having at least 50-60% identity, preferably about 60-70%, more preferably about 70-80%, or

about 80-90% identity with the first or second 7 transmembrane receptor profile of human C5RL (*e.g.*, has the recited degree of identity with amino acids 52-131 of SEQ ID NO:2 or with amino acids 153 to 291 of SEQ ID NO:2).

Preferred C5RL molecules of the present invention have an amino acid sequence sufficiently homologous or identical to the C5RL protein having the amino acid sequence of SEQ ID NO:2. As used herein, the term "sufficiently homologous" or "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains share at least about 50% identity, preferably 50-60% identity, more preferably 60-70%, and even more preferably 70-80%, 80-90%, 90-95% or more identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains, are defined herein as sufficiently homologous or identical. Furthermore, amino acid or nucleotide sequences which share at least about 50% identity, preferably 50-60% identity, more preferably 60-70%, and even more preferably 70-80%, 80-90%, 90-95% or more identity and share a common functional activity are defined herein as sufficiently homologous or identical.

As used interchangeably herein, a "C5RL activity", "biological activity of C5RL" or "functional activity of C5RL", refers to an activity exerted by a C5RL protein, polypeptide or nucleic acid molecule on a C5RL, for example, in a C5RL-expressing cell, as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a C5RL activity is a direct activity, such as an association with a C5RL-target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which a C5RL protein binds or interacts in nature, such that C5RL-mediated function is achieved. A C5RL target molecule can be a non-C5RL molecule, *i.e.*, a heterologous protein or polypeptide. In an exemplary embodiment, a C5RL target molecule is a C5RL ligand (*e.g.*, a complement protein fragment or anaphylatoxin, for example, a C5a polypeptide). In another exemplary embodiment, a C5RL target molecule is a heterotrimeric G protein. Alternatively, a C5RL activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the C5RL protein with a C5RL ligand.

In a preferred embodiment, a C5RL activity is at least one or more of the following activities: (i) interaction of a C5RL protein with soluble C5RL ligand; (ii) interaction of a C5RL protein with a membrane-bound or membrane-associated non-C5RL protein; (iii) interaction of a C5RL protein with an intracellular protein (*e.g.*, an intracellular enzyme or signal transduction molecule); and (iv) indirect interaction or activation of an intracellular protein (*e.g.*, a downstream signal transduction molecule).

In yet another preferred embodiment, a C5RL activity is at least one or more of the following activities: (1) modulation of cellular signal transduction, either *in vitro* or *in vivo*; (2) mediation of cell attachment and/or adhesion; (3) regulation of immune cell function; (4) regulation of cellular growth, differentiation or development (*e.g.*, immune cell growth, differentiation or development); (5) modulation of calcium mobilization (*e.g.*, modulation of calcium release from intracellular stores); (6) regulation of inflammatory processes; (7) modulation of cell host defense processes or mechanisms; and (8) modulation of smooth muscle cell contraction (*e.g.*, in the lung). C5RL proteins, nucleic acid molecules, fragments, antibodies and the like can accordingly be used to achieve any of the above described functions, for example in a cell or tissue expressing, producing or contacted with the protein, nucleic acid molecule, fragment, antibody, etc. Moreover, C5RL proteins and/or nucleic acid molecules can be used in screening assays as described herein to identify compounds that modulate any of the biological activities set forth above, in particular, pharmaceutical and/or therapeutic compounds useful in modulating (*e.g.*, inhibiting) immune cell function and/or inflammation, modulating (*e.g.*, inhibiting anaphylaxis), treating autoimmune disorders, asthma, arthritis (*e.g.*, rheumatoid arthritis) and/or inflammatory bowel disease, inhibiting tissue damage, and modulating immunological disorders, inflammation, allergy and host defense.

A preferred embodiment of the invention features isolated C5RL proteins and polypeptides having a C5RL activity, as described herein. Preferred C5RL proteins have at least one, preferably two, three, four, five or six transmembrane domains and a C5RL activity. In a preferred embodiment, a C5RL protein has a 7 transmembrane receptor profile and a C5RL activity. In another preferred embodiment, a C5RL protein has a 7 transmembrane receptor profile, a C5RL activity, and an amino acid sequence sufficiently homologous or identical to the amino acid sequence of SEQ ID NO:2.

A human C5RL cDNA was identified from a library derived from peripheral blood leukocyte (PBL) RNA, which is approximately 1374 nucleotides in length, and which encodes a protein which is approximately 337 amino acid residues in length. The human C5RL protein contains 7 transmembrane domains centered about around amino acids 50, 85, 120, 160, 215, 245 and 285 of SEQ ID NO:2 (*e.g.*, at about amino acids 31-61, 71-94, 109-130, 154-172, 205-226, 234-259 and 275-297 of SEQ ID NO:2). The human C5RL protein further contains a first 7 transmembrane receptor profile at about amino acids 52-131 of SEQ ID NO:2 and a second 7 transmembrane receptor profile at about amino acids 153-291 of SEQ ID NO:2.

The nucleotide sequence of the isolated human C5RL cDNA and the predicted amino acid sequence encoded by the C5RL cDNA are shown in Figures 1A and 1B, respectively and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the human C5RL cDNA was deposited with the American Type Culture Collection (ATCC), 10801

University Boulevard, Manassas, VA 20110-2209, on December 12, 2000 and assigned Accession Number PTA-2776. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit were made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human C5RL gene, which is approximately 1374 nucleotides in length, encodes a protein having a molecular weight of approximately 37 kD and which is approximately 337 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode C5RL proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify C5RL-encoding nucleic acids (*e.g.*, C5RL mRNA) and fragments for use as PCR primers for the amplification or mutation of C5RL nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated C5RL nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or 3, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a

portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776 as a hybridization probe, C5RL nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J. *et al., Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to C5RL nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1 or 3. The sequence of SEQ ID NO:1 corresponds to the human C5RL cDNA. This cDNA comprises sequences encoding the human C5RL protein (*i.e.*, "the coding region", from nucleotides 180-1190), a stop codon from nucleotides 1191-1193, as well as 5' untranslated sequences (nucleotides 1-179) and 3' untranslated sequences (nucleotides 1192 or 1194 to 1374). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 180-1190, corresponding to SEQ ID NO:3).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or 3, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776 is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 70-75%, preferably at least about 75-80%, more preferably at least about 80-85%, more preferably at least about 85-90%, and even more preferably at least about 90-95% or more identical to the nucleotide sequences shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a C5RL protein. The nucleotide sequence determined from the cloning of the C5RL genes allows for the generation of probes and primers designed for use in identifying and/or cloning other C5RL family members, as well as C5RL homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The probe/primer (*e.g.*, oligonucleotide) typically comprises about 5-10, 10-15, 15-20, 20-25 or more nucleotides that hybridize under stringent conditions to a sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, to an anti-sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or to a naturally occurring mutant of SEQ ID NO:1 or 3. Alternatively, the probe/primer (*e.g.*, oligonucleotide) comprises about 5-10, 10-15, 15-20, 20-25 or more consecutive nucleotides of a sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, of an anti-sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or of a naturally occurring mutant of SEQ ID NO:1 or 3.

Probes based on the C5RL nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a C5RL protein, such as by measuring a level of a C5RL-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting C5RL mRNA levels or determining whether a genomic C5RL gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a C5RL protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, which encodes a polypeptide having a C5RL biological activity (the biological activities of the C5RL proteins have previously been described), expressing the encoded portion of the C5RL protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the C5RL protein. In an exemplary embodiment, a nucleic acid molecule encoding a biologically active portion of a protein of the present invention comprises a nucleotide sequence which is greater that 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or 1300 nucleotides in length and hybridizes under stringent hybridization conditions to a complement of the nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, due to degeneracy of the genetic code and thus encode the same C5RL proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1 or 3. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2.

In addition to the C5RL nucleotide sequences shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the C5RL proteins may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the C5RL genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a C5RL protein, preferably a mammalian C5RL protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a C5RL gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in C5RL genes that are the result of natural allelic variation and that do not alter the functional activity of a C5RL protein are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding other C5RL family members and thus which have a nucleotide sequence which differs from the C5RL sequence of SEQ ID NO:1 or 3 are intended to be within the scope of the invention. For example, a primate C5RL cDNA can be identified based on the nucleotide sequence of human

C5RL. Nucleic acid molecules encoding C5RL proteins from different species, and thus which have a nucleotide sequence which differs from the C5RL sequences of SEQ ID NO:1 or 3 are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the C5RL cDNAs of the invention can be isolated based on their homology to the C5RL nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to a complement of the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776. In another embodiment, the nucleic acid is at least 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or 1300 nucleotides in length and hybridizes under stringent conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions includes hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions includes hybridization in 1X SSC, at about 65-70°C (or hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions includes hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45° C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate

to the above-recited values, *e.g.*, at 65-70°C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, $T_m(°C) = 2(\#\text{ of A + T bases}) + 4(\#\text{ of G + C bases})$. For hybrids between 18 and 49 base pairs in length, $T_m(°C) = 81.5 + 16.6(\log_{10}[Na^+]) + 0.41(\%G+C) - (600/N)$, where N is the number of bases in the hybrid, and $[Na^+]$ is the concentration of sodium ions in the hybridization buffer ($[Na^+]$ for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH_2PO_4, 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH_2PO_4, 1% SDS at 65°C, see *e.g.*, Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (or alternatively 0.2X SSC, 1% SDS).

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a complement of the sequence of SEQ ID NO:1 or 3 and corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural polypeptide).

In addition to naturally-occurring allelic variants of the C5RL sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, thereby leading to changes in the amino acid sequence of the encoded C5RL proteins, without altering the functional ability of the C5RL proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of C5RL (*e.g.*, the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue

is required for biological activity. For example, amino acid residues that are conserved among the C5RL proteins of the present invention and C5R1 proteins, are predicted to be particularly unamenable to alteration. Moreover, amino acid residues that are defined by the 7 transmembrane signature profiles are particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the C5RL proteins of the present invention and other members of the G protein coupled receptor protein family are not likely to be amenable to alteration (*e.g.*, the conserved residues depicted in Figure 2).

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding C5RL proteins that contain changes in amino acid residues that are not essential for activity. Such C5RL proteins differ in amino acid sequence from SEQ ID NO:2 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 70% identical to the amino acid sequence of SEQ ID NO:2. Preferably, the protein encoded by the nucleic acid molecule is at least about 70-75% identical to SEQ ID NO:2, more preferably at least about 75-80% identical to SEQ ID NO:2, more preferably at least about 80-85% identical to SEQ ID NO:2, even more preferably at least about 85-90% identical to SEQ ID NO:2, and even more preferably at least about 90-95% identical to SEQ ID NO:2.

An isolated nucleic acid molecule encoding a C5RL protein homologous to the protein of SEQ ID NO:2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1 or 3, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1 or 3 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a C5RL protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly

along all or part of a C5RL coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for C5RL biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or 3, or the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant C5RL protein can be assayed for the ability to (i) interact with a C5RL protein with soluble C5RL ligand; (ii) interact with a C5RL protein with a membrane-bound or membrane-associated non-C5RL protein; (iii) interact with a C5RL protein with an intracellular protein (*e.g.*, an intracellular enzyme or signal transduction molecule); and/or (iv) indirectly interact with or activate an intracellular protein (*e.g.*, a downstream signal transduction molecule).

In yet another preferred embodiment, a mutant C5RL protein can be assayed for the ability to (1) modulate cellular signal transduction, either *in vitro* or *in vivo*; (2) mediate cell attachment and/or adhesion; (3) regulate immune cell function; (4) regulate cellular growth, differentiation or development (*e.g.*, immune cell growth, differentiation or development); (5) modulate calcium mobilization (*e.g.*, modulate calcium release from intracellular stores); (6) regulate inflammatory processes; (7) modulate cell host defense processes or mechanisms; and/or (8) modulate smooth muscle cell contraction (*e.g.*, in the lung).

In addition to the nucleic acid molecules encoding C5RL proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire C5RL coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding C5RL. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human C5RL corresponds to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding C5RL. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding C5RL disclosed herein (*e.g.*, SEQ ID NO:3), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be

complementary to the entire coding region of C5RL mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of C5RL mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of C5RL mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a C5RL protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified

to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haseloff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave C5RL mRNA transcripts to thereby inhibit translation of C5RL mRNA. A ribozyme having specificity for a C5RL-encoding nucleic acid can be designed based upon the nucleotide sequence of a C5RL cDNA disclosed herein (*i.e.*, SEQ ID NO:1 or 3). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a C5RL-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, C5RL mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, C5RL gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the C5RL (*e.g.*, the C5RL promoter and/or enhancers) to form triple helical structures that prevent transcription of the C5RL gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioessays* 14(12):807-15.

In yet another embodiment, the C5RL nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup, B. and Nielsen, P.E. (1996) *Bioorg. Med. Chem.* 4(1):5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup and Nielsen (1996) *supra* and Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670-675.

PNAs of C5RL nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of C5RL nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes (*e.g.*, S1 nucleases (Hyrup and Nielsen (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup and Nielsen (1996) *supra*; Perry-O'Keefe *et al.* (1996) *supra*).

In another embodiment, PNAs of C5RL can be modified (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of C5RL nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNAse H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup and Nielsen (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup and Nielsen (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acids Res.* 17:5973-88). PNA monomers are then

coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810, published December 15, 1988) or the blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134, published April 25, 1988). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.* (1988) *Biotechniques* 6:958-976) or intercalating agents (see, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

Furthermore, given the fact that an important use for the C5RL molecules of the present invention is in the screening for C5RL ligands (*e.g.*, surrogate ligands) and/or C5RL modulators, it is intended that the following are also within the scope of the present invention: isolated nucleic acids which encode and C5RL ligands or C5RL modulators, probes and/or primers useful for identifying C5RL ligands or C5RL modulators based on the sequences of nucleic acids which encode and C5RL ligands or C5RL modulators, isolated nucleic acid molecules which are complementary or antisense to the sequences of nucleic acids which encode and C5RL ligands or C5RL modulators, isolated nucleic acid molecules which are at least about 60-65%, preferably at least about 70-75%, more preferable at least about 80-85%, and even more preferably at least about 90-95% or more identical to the sequences of nucleic acids which encode and C5RL ligands or C5RL modulators, portions of nucleic acids which encode and C5RL ligands or C5RL modulators (*e.g.*, biologically-active portions), naturally-occurring allelic variants of nucleic acids which encode and C5RL ligands or C5RL modulators, nucleic acid molecules which hybridize under stringent hybridization conditions to nucleic acids which encode and C5RL ligands or C5RL modulators, functionally-active mutants of nucleic acids which encode and C5RL ligands or C5RL modulators, PNAs of nucleic acids which encode and C5RL ligands or C5RL modulators, as well as vectors containing a nucleic acid encoding a C5RL ligand or C5RL modulator, described herein, host cells into which an expression vector encoding a C5RL ligand or C5RL modulator has been introduced, and homologous recombinant animal which express C5RL ligands or C5RL modulators.

II. Isolated C5RL Proteins and Anti-C5RL Antibodies

One aspect of the invention pertains to isolated C5RL proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-C5RL antibodies. In one embodiment, native C5RL proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, C5RL proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a C5RL protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the C5RL protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of C5RL protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of C5RL protein having less than about 30% (by dry weight) of non-C5RL protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-C5RL protein, still more preferably less than about 10% of non-C5RL protein, and most preferably less than about 5% non-C5RL protein. When the C5RL protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of C5RL protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of C5RL protein having less than about 30% (by dry weight) of chemical precursors or non-C5RL chemicals, more preferably less than about 20% chemical precursors or non-C5RL chemicals, still more preferably less than about 10% chemical precursors or non-C5RL chemicals, and most preferably less than about 5% chemical precursors or non-C5RL chemicals.

Biologically active portions of a C5RL protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the C5RL protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length C5RL proteins, and exhibit at least one activity of a C5RL protein. Typically, biologically active portions comprise a

domain or motif with at least one activity of the C5RL protein. A biologically active portion of a C5RL protein can be a polypeptide which is, for example, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 375, 300, 325 or more amino acids in length.

In one embodiment, a biologically active portion of a C5RL protein comprises at least one, preferably two, more preferably three, four, five, six or seven transmembrane domains. In another embodiment, a biologically active portion of a C5RL protein comprises at least one, preferably two 7 transmembrane receptor profiles or domains.

It is to be understood that a preferred biologically active portion of a C5RL protein of the present invention may contain at least one of the above-identified structural domains and/or profiles. A more preferred biologically active portion of a C5RL protein may contain at least two of the above-identified structural domains and/or profiles. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native C5RL protein.

In a preferred embodiment, the C5RL protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the C5RL protein is substantially homologous or identical to SEQ ID NO:2, and retains the functional activity of the protein of SEQ ID NO:2, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the C5RL protein is a protein which comprises an amino acid sequence at least about 75% identical to the amino acid sequence of SEQ ID NO:2 and retains the functional activity of the C5RL proteins of SEQ ID NO:2, respectively. Preferably, the protein is at least about 70-75% identical to SEQ ID NO:2, more preferably at least about 75-80% identical to SEQ ID NO:2, more preferably at least about 80-85% identical to SEQ ID NO:2, even more preferably at least about 85-90% identical to SEQ ID NO:2, and most preferably at least about 90-95% or more identical to SEQ ID NO:2.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the C5RL amino acid sequence of SEQ ID NO:2 having 337 amino acid residues, at least 101, preferably at least 135, more preferably at least 169, even more preferably at least 202, and even more preferably at least 236, 270 or 303 amino

acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. In a particularly preferred embodiment, percent identity is determined over the entire length of the sequence being compared, *i.e.*, the sequences are globally aligned for comparison purposed.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at Genetics Computer Group Web Server), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Genetics Computer Group Web Server), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A preferred, non-limiting example of parameters to be used in conjunction with the GAP program include a Blosum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of Meyers, E. and Miller, W. (*Comput. Appl. Biosci.* 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0 or version 2.0U), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and polypeptide sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the BLASTN and BLASTX programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to C5RL nucleic acid molecules of the invention. BLAST protein searches can be performed with the BLASTX program, score = 100, wordlength = 3, and a

Blosum62 matrix to obtain amino acid sequences homologous to C5RL polypeptide molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, BLASTX and BLASTN) can be used. See National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health Web Server.

Proteins and protein fragments of the present invention include proteins with amino acid sequence lengths that are at least 25% (more preferably at least 50%, and even more preferably at least 75%) of the length of a disclosed protein and have at least 60-70% sequence identity (more preferably, at least 70-75% identity and even more preferably at least 75-80%, 80-85%, 85-90%, 90-95% or more identity) with that disclosed protein, where sequence identity is determined as described herein.

The invention also provides C5RL chimeric or fusion proteins. As used herein, a C5RL "chimeric protein" or "fusion protein" comprises a C5RL polypeptide operatively linked to a non-C5RL polypeptide. A "C5RL polypeptide" refers to a polypeptide having an amino acid sequence corresponding to C5RL, whereas a "non-C5RL polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the C5RL protein, *e.g.*, a protein which is different from the C5RL protein and which is derived from the same or a different organism. Within a C5RL fusion protein the C5RL polypeptide can correspond to all or a portion of a C5RL protein. In a preferred embodiment, a C5RL fusion protein comprises at least one biologically active portion of a C5RL protein. In another preferred embodiment, a C5RL fusion protein comprises at least two biologically active portions of a C5RL protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the C5RL polypeptide and the non-C5RL polypeptide are fused in-frame to each other. The non-C5RL polypeptide can be fused to the N-terminus or C-terminus of the C5RL polypeptide.

For example, in one embodiment, the fusion protein is a GST-C5RL fusion protein in which the C5RL sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant C5RL. In another embodiment, the fusion protein is a C5RL protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of C5RL can be increased through use of a heterologous signal sequence.

The C5RL fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The C5RL fusion proteins can be used to affect the bioavailability of a C5RL substrate. Use of C5RL

fusion proteins may be useful therapeutically for the treatment of immunological disorders. Moreover, the C5RL-fusion proteins of the invention can be used as immunogens to produce anti-C5RL antibodies in a subject, to purify C5RL ligands and in screening assays to identify molecules which inhibit the interaction of C5RL with a C5RL ligand.

Preferably, a C5RL chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons:1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A C5RL-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the C5RL protein.

The present invention also pertains to variants of the C5RL proteins which function as either C5RL agonists (mimetics) or as C5RL antagonists. Variants of the C5RL proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a C5RL protein. An agonist of the C5RL proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a C5RL protein. An antagonist of a C5RL protein can inhibit one or more of the activities of the naturally occurring form of the C5RL protein by, for example, competitively inhibiting the protease activity of a C5RL protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the C5RL protein.

In one embodiment, variants of a C5RL protein which function as either C5RL agonists (mimetics) or as C5RL antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a C5RL protein for C5RL protein agonist or antagonist activity. In one embodiment, a variegated library of C5RL variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a

variegated gene library. A variegated library of C5RL variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential C5RL sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of C5RL sequences therein. There are a variety of methods which can be used to produce libraries of potential C5RL variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential C5RL sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of a C5RL protein coding sequence can be used to generate a variegated population of C5RL fragments for screening and subsequent selection of variants of a C5RL protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a C5RL coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the C5RL protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of C5RL proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify C5RL variants (Arkin and Youvan

(1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delagrave *et al.* (1993) *Protein Eng.* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated C5RL library. For example, a library of expression vectors can be transfected into a cell line which ordinarily synthesizes C5RL. The transfected cells are then cultured such that a particular mutant C5RL is expressed and the effect of expression of the mutant on C5RL activity in the cell can be detected, *e.g.*, by any of a number of activity assays for native C5RL protein. Plasmid DNA can then be recovered from the cells which score for modulated C5RL activity, and the individual clones further characterized.

An isolated C5RL protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind C5RL using standard techniques for polyclonal and monoclonal antibody preparation. A full-length C5RL protein can be used or, alternatively, the invention provides antigenic peptide fragments of C5RL for use as immunogens. The antigenic peptide of C5RL comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of C5RL such that an antibody raised against the peptide forms a specific immune complex with C5RL. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of C5RL that are located within the extracellular and/or cytoplasmic portions of the C5RL protein (*e.g.*, within the extracellular and/or cytoplasmic loops or termini).

A C5RL immunogen typically is used to prepare antibodies by immunizing a suitable subject (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed C5RL protein or a chemically synthesized C5RL polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic C5RL preparation induces a polyclonal anti-C5RL antibody response.

Accordingly, another aspect of the invention pertains to anti-C5RL antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as C5RL. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind C5RL. The term "monoclonal antibody" or

"monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of C5RL. A monoclonal antibody composition thus typically displays a single binding affinity for a particular C5RL protein with which it immunoreacts.

Polyclonal anti-C5RL antibodies can be prepared as described above by immunizing a suitable subject with a C5RL immunogen. The anti-C5RL antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized C5RL. If desired, the antibody molecules directed against C5RL can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-C5RL antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem.* 255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985) *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally Kenneth, R.H. in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); Lerner, E.A. (1981) *Yale J. Biol. Med.* 54:387-402; Gefter, M.L. *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a C5RL immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds C5RL.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-C5RL monoclonal antibody (see, *e.g.*, Galfre, G. *et al.* (1977) *Nature* 266:55052; Gefter *et al.* (1977) *supra*; Lerner (1981) *supra*; Kenneth (1980) *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line.

Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind C5RL, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-C5RL antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with C5RL to thereby isolate immunoglobulin library members that bind C5RL. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication No. WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication No. WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1369-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clackson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nucleic Acids Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* (1990) *Nature* 348:552-554.

Additionally, recombinant anti-C5RL antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in

Robinson *et al.* International Application No. PCT/US86/02269; Akira *et al.* European Patent Application No. 184,187; Taniguchi, M., European Patent Application No. 171,496; Morrison *et al.* European Patent Application No. 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application No. 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S.L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *Biotechniques* 4:214; Winter U.S. Patent No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyen *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-C5RL antibody (*e.g.*, monoclonal antibody) can be used to isolate C5RL by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-C5RL antibody can facilitate the purification of natural C5RL from cells and of recombinantly produced C5RL expressed in host cells. Moreover, an anti-C5RL antibody can be used to detect C5RL protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the C5RL protein. Anti-C5RL antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Furthermore, given the fact that an important use for the C5RL molecules of the present invention is in the screening for C5RL ligands (*e.g.*, surrogate ligands) and/or C5RL modulators, it is intended that the following are also within the scope of the present invention: "isolated" or "purified" C5RL ligands or C5RL modulators, biologically-active portions of C5RL ligands or C5RL modulators, chimeric or fusion

proteins comprising all or a portion of a C5RL ligand or C5RL modulator, and antibodies comprising all or a portion of a C5RL ligand or C5RL modulator.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a C5RL protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel (1990) *Methods Enzymol.* 185:3-7. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those

skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, C5RL proteins, mutant forms of C5RL proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of C5RL proteins in prokaryotic or eukaryotic cells. For example, C5RL proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel (1990) *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in C5RL activity assays (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for C5RL proteins, for example. In a preferred embodiment, a C5RL fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.* (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.* (1990) *Methods Enzymol.* 185:60-89). Target gene expression from the pTrc vector relies on

host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S. (1990) *Methods Enzymol.* 185:119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the C5RL expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari *et al.* (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.* (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).

Alternatively, C5RL proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.*

(1987) *Genes Dev.* 1:268-277); lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to C5RL mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a C5RL protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells *via* conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a C5RL protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a C5RL protein. Accordingly, the invention further provides methods for producing a C5RL protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a C5RL protein has been introduced) in a suitable medium such that a C5RL protein is produced. In another embodiment, the method further comprises isolating a C5RL protein from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which C5RL-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous C5RL sequences have been introduced into their genome or homologous recombinant animals in which endogenous C5RL sequences have been altered. Such animals are useful for studying the function and/or activity of a C5RL and for identifying and/or evaluating modulators of C5RL activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a

transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous C5RL gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a C5RL-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The C5RL cDNA sequence, *e.g.*, that of SEQ ID NO:1 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman orthologue of a human C5RL gene, such as a mouse or rat C5RL gene, can be used as a transgene. Alternatively, a C5RL gene homologue can be isolated based on hybridization to the C5RL cDNA sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776 (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a C5RL transgene to direct expression of a C5RL protein to particular cells. Methods for generating transgenic animals *via* embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a C5RL transgene in its genome and/or expression of C5RL mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a C5RL protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a C5RL gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the C5RL gene. The C5RL gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:1 or 3), but more preferably, is a non-human homologue of a human C5RL gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1 or 3). For example, a mouse C5RL gene can be used to construct a homologous recombination vector suitable for altering an endogenous C5RL gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous C5RL gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous C5RL gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous C5RL protein). In the homologous recombination vector, the altered portion of the C5RL gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the C5RL gene to allow for homologous recombination to occur between the exogenous C5RL gene carried by the vector and an endogenous C5RL gene in an embryonic stem cell. The additional flanking C5RL nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see *e.g.*, Thomas, K.R. and Capecchi, M.R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced C5RL gene has homologously recombined with the endogenous C5RL gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, E.J., ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The C5RL nucleic acid molecules, C5RL proteins, anti-C5RL antibodies, C5RL ligands, and C5RL modulators (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, antibody, or modulatory compound and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a C5RL protein or anti-C5RL antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by

incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate,

polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent No. 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, a C5RL protein of the invention has one or more of the following activities: (i) interaction of a C5RL protein with soluble C5RL ligand; (ii) interaction of a C5RL protein with a membrane-bound or membrane-associated non-C5RL protein; (iii) interaction of a C5RL protein with an intracellular protein (*e.g.*, an intracellular enzyme or signal transduction molecule); and (iv) indirect interaction or activation of an intracellular protein (*e.g.*, a downstream signal transduction molecule, and can thus be used in, for example, (1) modulation of cellular signal transduction, either *in vitro* or *in vivo*; (2) mediation of cell attachment and/or adhesion; (3) regulation of immune cell function; (4) regulation of cellular growth, differentiation or development (*e.g.*, immune cell growth, differentiation or development); (5) modulation of calcium mobilization (*e.g.*, modulation of calcium release from intracellular stores); (6) regulation of inflammatory processes; (7) modulation of cell host defense processes or mechanisms; and (8) modulation of smooth muscle cell contraction (*e.g.*, in the lung). The isolated nucleic acid molecules of the invention can be used, for example, to express C5RL protein (*e.g.*, *via* a recombinant expression vector in a host cell in gene therapy applications), to detect C5RL mRNA (*e.g.*, in a biological sample) or a genetic alteration in a C5RL gene, and to modulate C5RL activity, as described further below. The C5RL proteins can be used to treat disorders characterized by insufficient or excessive production of a C5RL protein and/or C5RL ligand. In addition, the C5RL proteins can be used to screen drugs or compounds which modulate the C5RL activity as well as to

treat disorders characterized by insufficient or excessive production of C5RL protein or production of C5RL protein forms which have decreased or aberrant activity compared to C5RL wild type protein. Moreover, the anti-C5RL antibodies of the invention can be used to detect and isolate C5RL proteins, regulate the bioavailability of C5RL proteins, and modulate C5RL activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to C5RL proteins, or have a stimulatory or inhibitory effect on, for example, C5RL expression or C5RL activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a C5RL protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a C5RL protein on the cell surface is contacted with a test compound and the ability of the test compound to bind to the C5RL protein determined. The cell, for

example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to a C5RL protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the C5RL protein can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a test compound to interact with a C5RL protein without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test compound with a C5RL protein without the labeling of either the test compound or the receptor. McConnell, H.M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and receptor.

In a preferred embodiment, the assay comprises contacting a cell which expresses a C5RL protein or biologically active portion thereof, on the cell surface with a C5RL ligand, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the C5RL protein or biologically active portion thereof, wherein determining the ability of the test compound to interact with the C5RL protein or biologically active portion thereof, comprises determining the ability of the test compound to preferentially bind to the C5RL protein or biologically active portion thereof, as compared to the ability of the C5RL ligand to bind to the C5RL protein or biologically active portion thereof.

Determining the ability of the C5RL ligand or C5RL modulator to bind to or interact with a C5RL protein or biologically active portion thereof, can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the C5RL ligand or modulator to bind to or interact with a C5RL protein or biologically active portion thereof, can be accomplished by determining the activity of a C5RL protein or of a downstream C5RL target molecule. For example, the target molecule can be a cellular second messenger, and the activity of the target molecule can be determined by detecting induction of the target (*i.e.*, intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene

(comprising a C5RL-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, a proliferative response or an inflammatory response. Accordingly, in one embodiment the present invention involves a method of identifying a compound which modulates the activity of a C5RL protein, comprising contacting a cell which expresses a C5RL protein with a test compound, determining the ability of the test compound to modulate the activity the C5RL protein, and identifying the compound as a modulator of C5RL activity. In another embodiment, the present invention involves a method of identifying a compound which modulates the activity of a C5RL protein, comprising contacting a cell which expresses a C5RL protein with a test compound, determining the ability of the test compound to modulate the activity of a downstream C5RL target molecule, and identifying the compound as a modulator of C5RL activity.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a C5RL protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the C5RL protein or biologically active portion thereof is determined. Binding of the test compound to the C5RL protein can be determined either directly or indirectly as described above. Binding of the test compound to the C5RL protein can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore™). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In a preferred embodiment, the assay includes contacting the C5RL protein or biologically active portion thereof with a known ligand which binds C5RL to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a C5RL protein, wherein determining the ability of the test compound to interact with a C5RL protein comprises determining the ability of the test compound to preferentially bind to C5RL or biologically active portion thereof as compared to the known ligand.

In another embodiment, the assay is a cell-free assay in which a C5RL protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the C5RL protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a C5RL protein can be accomplished, for example, by determining the ability of the C5RL protein to modulate the activity of a downstream C5RL target molecule by one of the methods described above for cell-based

assays. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a C5RL protein or biologically active portion thereof with a known ligand which binds the C5RL protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the C5RL protein, wherein determining the ability of the test compound to interact with the C5RL protein comprises determining the ability of the test compound to preferentially bind to or modulate the activity of a C5RL target molecule, as compared to the known ligand.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (*e.g.*, C5RL proteins or biologically active portions thereof or C5RL proteins). In the case of cell-free assays in which a membrane-bound form an isolated protein is used (*e.g.*, a C5RL protein) it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either C5RL or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a C5RL protein, or interaction of a C5RL protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ C5RL fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or C5RL protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either

directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of C5RL binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a C5RL protein or a C5RL target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated C5RL protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with C5RL protein or target molecules but which do not interfere with binding of the C5RL protein to its target molecule can be derivatized to the wells of the plate, and unbound target or C5RL protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the C5RL protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the C5RL protein or target molecule.

In another embodiment, modulators of C5RL expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of C5RL mRNA or protein in the cell is determined. The level of expression of C5RL mRNA or protein in the presence of the candidate compound is compared to the level of expression of C5RL mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of C5RL expression based on this comparison. For example, when expression of C5RL mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of C5RL mRNA or protein expression. Alternatively, when expression of C5RL mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of C5RL mRNA or protein expression. The level of C5RL mRNA or protein expression in the cells can be determined by methods described herein for detecting C5RL mRNA or protein.

In yet another aspect of the invention, the C5RL proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with C5RL ("C5RL-binding proteins" or "C5RL-bp") and are involved in C5RL activity. Such C5RL-binding proteins are also likely to be involved in the

propagation of signals by the C5RL proteins as, for example, downstream elements of a C5RL-mediated signaling pathway. Alternatively, such C5RL-binding proteins are likely to be cell-surface molecules associated with non-C5RL expressing cells, wherein such C5RL-binding proteins are involved in chemoattraction.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a C5RL protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a C5RL-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the C5RL protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a C5RL modulating agent, an antisense C5RL nucleic acid molecule, a C5RL-specific antibody, or a C5RL-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the C5RL nucleotide sequences, described herein, can be used to map the location of the C5RL genes on a chromosome. The mapping of the C5RL sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, C5RL genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the C5RL nucleotide sequences. Computer analysis of the C5RL sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the C5RL sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the C5RL nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a 9o, 1p, or 1v sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been

blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data (such data are found, for example, in McKusick, V., Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the C5RL gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The C5RL sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is

digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the C5RL nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The C5RL nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from C5RL nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial C5RL Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator

of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.*, another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 or 3 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the C5RL nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1, having a length of at least 20 bases, preferably at least 30 bases.

The C5RL nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such C5RL probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, C5RL primers or probes can be used to screen tissue culture for contamination (*i.e.*, screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining C5RL protein and/or nucleic acid expression as well as C5RL activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant C5RL expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with C5RL protein, nucleic acid expression or

activity. For example, mutations in a C5RL gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with C5RL protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of C5RL in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of C5RL protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting C5RL protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes C5RL protein such that the presence of C5RL protein or nucleic acid is detected in the biological sample. A preferred agent for detecting C5RL mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to C5RL mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length C5RL nucleic acid, such as the nucleic acid of SEQ ID NO:1 or 3, or a fragment or portion of a C5RL nucleic acid such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to C5RL mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting C5RL protein is an antibody capable of binding to C5RL protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect C5RL mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of C5RL mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of C5RL protein include enzyme linked

immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of C5RL genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of C5RL protein include introducing into a subject a labeled anti-C5RL antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting C5RL protein, mRNA, or genomic DNA, such that the presence of C5RL protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of C5RL protein, mRNA or genomic DNA in the control sample with the presence of C5RL protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of C5RL in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting C5RL protein or mRNA in a biological sample; means for determining the amount of C5RL in the sample; and means for comparing the amount of C5RL in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect C5RL protein or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant C5RL expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with C5RL protein, nucleic acid expression or activity such as an inflammatory disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing an inflammatory disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant C5RL expression or activity in which a test sample is obtained from a subject and C5RL protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of C5RL protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated

with aberrant C5RL expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant C5RL expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as an inflammatory disorder. Alternatively, such methods can be used to determine whether a subject can be effectively treated with an agent for an inflammatory disease. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant C5RL expression or activity in which a test sample is obtained and C5RL protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of C5RL protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant C5RL expression or activity.)

The methods of the invention can also be used to detect genetic alterations in a C5RL gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by an aberrant inflammatory response. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a C5RL-protein, or the mis-expression of the C5RL gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a C5RL gene; 2) an addition of one or more nucleotides to a C5RL gene; 3) a substitution of one or more nucleotides of a C5RL gene, 4) a chromosomal rearrangement of a C5RL gene; 5) an alteration in the level of a messenger RNA transcript of a C5RL gene, 6) aberrant modification of a C5RL gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a C5RL gene, 8) a non-wild type level of a C5RL-protein, 9) allelic loss of a C5RL gene, and 10) inappropriate post-translational modification of a C5RL-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a C5RL gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a

ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the C5RL-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a C5RL gene under conditions such that hybridization and amplification of the C5RL-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a C5RL gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in C5RL can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Hum. Mutat.* 7:244-255; Kozal, M.J. *et al.* (1996) *Nat. Med.* 2:753-759). For example, genetic mutations in C5RL can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.* (1996) *supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the

identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the C5RL gene and detect mutations by comparing the sequence of the sample C5RL with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the C5RL gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type C5RL sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in C5RL cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase

from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a C5RL sequence, *e.g.*, a wild-type C5RL sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in C5RL genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control C5RL nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a C5RL gene.

Furthermore, any cell type or tissue in which C5RL is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of a C5RL protein (*e.g.*, modulation of an inflammatory response) an be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase C5RL gene expression, protein levels, or upregulate C5RL activity, can be monitored in clinical trails of subjects exhibiting decreased C5RL gene expression, protein levels, or downregulated C5RL activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease C5RL gene expression, protein levels, or downregulate C5RL activity, can be monitored in clinical trails of subjects exhibiting increased C5RL gene expression, protein levels, or upregulated C5RL activity. In such clinical trials, the expression or activity of a C5RL gene, and preferably, other genes that have been implicated in, for example, an inflammatory disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including C5RL, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates C5RL activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on inflammatory disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of C5RL and other genes implicated in the inflammatory disorder, respectively. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of C5RL or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a C5RL protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the C5RL protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the C5RL protein, mRNA, or genomic DNA in the pre-administration sample with the C5RL protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of C5RL to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of C5RL to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent. According to such an embodiment, C5RL expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant C5RL expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or

modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the C5RL molecules of the present invention or C5RL modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant C5RL expression or activity, by administering to the subject a C5RL or an agent which modulates C5RL expression or at least one C5RL activity. Subjects at risk for a disease which is caused or contributed to by aberrant C5RL expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the C5RL aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of C5RL aberrancy, for example, a C5RL, C5RL agonist or C5RL antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the present invention are further discussed in the following subsections.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating C5RL expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a C5RL molecule of the present invention such that the activity of a C5RL is modulated. Alternatively, the modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of C5RL protein activity associated with the cell. An agent that modulates C5RL protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a C5RL protein (*e.g.*, a carbohydrate), a C5RL antibody, a C5RL agonist or antagonist, a peptidomimetic of a C5RL agonist or antagonist, or other small molecule.

In one embodiment, the agent stimulates one or more C5RL activities. Examples of such stimulatory agents include active C5RL protein and a nucleic acid molecule encoding C5RL that has been introduced into the cell. In another embodiment, the agent inhibits one or more C5RL activities. Examples of such inhibitory agents include antisense C5RL nucleic acid molecules and anti-C5RL antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a C5RL protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) C5RL expression or activity. In another embodiment, the method involves administering a C5RL protein or nucleic acid molecule as therapy to compensate for reduced or aberrant C5RL expression or activity.

Stimulation of C5RL activity is desirable in situations in which C5RL is abnormally downregulated and/or in which increased C5RL activity is likely to have a beneficial effect. Likewise, inhibition of C5RL activity is desirable in situations in which C5RL is abnormally upregulated and/or in which decreased C5RL activity is likely to have a beneficial effect (*e.g.*, inflammation).

3. Pharmacogenomics

The C5RL molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on C5RL activity (*e.g.*, C5RL gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, inflammatory disorders) associated with aberrant C5RL activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a C5RL molecule or C5RL modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a C5RL molecule or C5RL modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, M. (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. (1997) *Clin. Chem.* 43(2):254-266. In general, two types

of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drugs target is known (*e.g.*, a C5RL protein or C5RL protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug.

These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a C5RL molecule or C5RL modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a C5RL molecule or C5RL modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and the Sequence Listing, are hereby incorporated by reference. References throughout the instant specification to websites and web servers also contain publicly-available information which can be accessed electronically through the World Wide Web.

EXAMPLES

Example 1: Identification and Characterization of a C5RL cDNA

In this example, the identification and characterization of the genes encoding human C5RL (also referred to as "L265B_P0091B07") is described.

Isolation of the human C5RL cDNAs

A clone designated L265B_P0091B07 was identified in a cDNA library derived from human peripheral blood leukocyte (PBL) RNA. The L265B_P0091B07 cDNA clone is in the pED6dpc2 vector which was derived from pED6dpc1 by insertion of a polylinker to facilitate cDNA cloning (Kaufman *et al.* (1991) *Nucleic Acids Res.* 19:4485-4490). The human clone contained an insert of approximately 1374 bp containing a protein-encoding sequence (*i.e.*, an open reading frame) of approximately 1011 nucleotides capable of encoding approximately 337 amino acids of C5RL.

The nucleotide sequence encoding the human C5RL protein is shown in Figure 1A and is set forth as SEQ ID NO:1. The full length protein encoded by this nucleic acid includes about 337 amino acids and has the amino acid sequence shown in Figure 1B and set forth as SEQ ID NO:2. The coding portion (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3. Clone L265B_P0091B07, comprising the coding region of human C5RL, was deposited under the name "L265_P0091B07" with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on December 12, 2000, and assigned Accession No. PTA-2776.

Analysis of Human C5RL

A BLASTN™ search (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403) using the nucleotide sequence of human C5RL revealed that C5RL is significantly similar to a C5a anaphylatoxin chemotactic receptor family member (which belongs to family 1 GPCRs) orthologs having been isolated from rabbit (GenBank™ Accession No. AF068680), guinea pig (SwissProt™ Accession No. O70129), Pongo pygmaeus (GenBank™ Accession No. X97732), gorilla (GenBank™ Accession No. X97733), and rat (GenBank™ Accession No. AB003042). Human C5RL is also similar to a previously identified human C5a receptor termed C5R1 which was identified in differentiated HL60 cells (Boulay *et al.* (1992) *Biochemistry* 30:2993-2999, GenBank™ Accession No. M62505).

The C5RL protein is predicted to have seven transmembrane domains (TopPred predicated) centered at about amino acids 50, 85, 120, 160, 215, 245 and 285 of SEQ ID NO:2. Boundaries of these transmembrane domains, predicted by alignment with corresponding transmembrane domains for related proteins, for example, guinea pig C5R1 and human C5R1 are located at about amino acids 31-61, 71-94, 109-130, 154-172, 205-226, 234-259 and 275-297 of SEQ ID NO:2. A Hidden Markov Model search also resulted in the identification of at least two 7_transmembrane receptor profiles at about amino acids 52-131 and 153-291 of SEQ ID NO:2.

C5RL is also predicted to contain the following sites: an Asn glycosylation site at amino acids 3-6; a cAMP-dependent protein kinase phosphorylation site at amino acids 324-327; protein kinase C phosphorylation sites at amino acids 17-19 and at amino acids 323-325; casein kinase II phosphorylation sites at amino acids 194-197, at amino acids 327-330 and at amino acids 333-336; and N-myristoylation sites at amino acids 26-31, amino acids 49-54, amino acids 103-108, amino acids 150-155, amino acids 156-151, at amino acids 191-196, at amino acids 192-197, at amino acids 253-258, at amino acids 278-283 and at amino acids 316-321.

Tissue Distribution of C5RL mRNA

To assess tissue distribution of the human C5RL gene, Northern analyses were performed as follows. Clontech Human Multiple Tissue Northern (catalog no. 7780-1, user manual PT1200-1) and Human Multiple Tissue Expression Array (catalog no. 7775-1, user manual PT3307-1) membranes were probed using the manufacturer's protocols and ^{32}P-labeled DNA probes. Probes containing L265B_P0091B07-specific sequence (nt 1-763 SEQ NO:1) were generated as follows. Plasmid DNA from an isolated L265B_P0091B07 colony was prepared using the QIAprep™ Spin Miniprep Kit and protocol. Subsequently, the DNA was restriction digested with EcoRI and PstI. Restriction fragments were size-fractionated by gel electrophoresis on 1.5% agarose, 0.1μg/ml ethidium bromide, 1x TAE gels (Maniatis *et al.*, 1982). The ethidium bromide-stained DNA band of the appropriate size (~760bp) was excised from the agarose gel. Next, the DNA was extracted from the agarose using the Clontech NucleoSpin™ Nucleic Acid Purification Kit and manufacturer's protocol. The extracted DNA was labeled with Redivue™ (α-^{32}P)dCTP (Amersham Pharmacia) using the Prime-It II™ Random Primer Labeling Kit and protocol (Stratagene). Unincorporated (alpha ^{32}P)dCTP was removed with Amersham's NICK™ column and protocol.

Human C5RL (L265B_P0091B07) transcript is predominately expressed in spleen, peripheral blood leukocytes, bone marrow, and lung. Notably, expression of C5RL is predominant in immunological cells and/or tissues and in lung.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 70% identical to the nucleotide sequence of SEQ ID NO:1 or 3, the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

2. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 80% identical to the nucleotide sequence of SEQ ID NO:1 or 3, the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

3. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 90% identical to the nucleotide sequence of SEQ ID NO:1 or 3, the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

4. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:1 or 3, the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

5. An isolated nucleic acid molecule comprising a fragment of at least 30 nucleotides of the nucleotide sequence of SEQ ID NO:3, or a complement thereof.

6. An isolated nucleic acid molecule comprising a fragment of at least 200 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

7. An isolated nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 70% identical to the amino acid sequence of SEQ ID NO:2 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

8. An isolated nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 80% identical to the amino acid sequence of SEQ ID NO:2 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

9. An isolated nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 90% identical to the amino acid sequence of SEQ ID NO:2 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

10. An isolated nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 95% identical to the amino acid sequence of SEQ ID NO:2 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

11. An isolated nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2.

12. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 or 3 under stringent conditions.

13. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, or a complement thereof.

14. An isolated nucleic acid molecule comprising a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.

15. The nucleic acid molecule of any one of claims 1-14, further comprising vector nucleic acid sequences.

16. The nucleic acid molecule of any one of claims 1-14, further comprising nucleic acid sequences encoding a heterologous polypeptide.

17. A host cell which contains the nucleic acid molecule of any one of claims 1-14.

18. The host cell of claim 17 which is a mammalian host cell.

19. A non-human mammalian host cell containing the nucleic acid molecule of any one of claims 1-14.

20. An isolated polypeptide comprising a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:2.

21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 or 3 under stringent conditions.

22. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 70% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 80% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

24. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 90% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

25. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

26. An isolated polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2 or the polypeptide encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

27. An isolated polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence of SEQ ID NO:2 or the polypeptide encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

28. An isolated polypeptide comprising an amino acid sequence which is at least 90% identical to the amino acid sequence of SEQ ID NO:2 or the polypeptide encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

29. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:2 or the polypeptide encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

30. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776.

31. The polypeptide of any one of claims 20-30 further comprising heterologous amino acid sequences.

32. An antibody which selectively binds to a polypeptide of any one of claims 20-30.

33. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number PTA-2776, comprising culturing the host cell of claim 17 under conditions in which the nucleic acid molecule is expressed.

34. A method for producing a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:2, comprising culturing the host cell of claim 17 under conditions in which the nucleic acid molecule is expressed.

35. A method for producing a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to complement of a nucleic acid molecule comprising SEQ ID NO:1 or 3 under stringent conditions, comprising culturing the host cell of claim 17 under conditions in which the nucleic acid molecule is expressed.

36. A method for detecting the presence of a polypeptide of any one of claims 20-30 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of any one of claims 20-30 in the sample.

37. The method of claim 36, wherein the compound which binds to the polypeptide is an antibody.

38. A kit comprising a compound which selectively binds to a polypeptide of any one of claims 20-30 and instructions for use.

39. A method for detecting the presence of a nucleic acid molecule of any one of claims 1-14 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any one of claims 1-14 in the sample.

40. The method of claim 39, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

41. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1-14 and instructions for use.

42. A method for identifying a compound which binds to a polypeptide of any one of claims 20-30 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

43. The method of claim 42, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for C5RL activity.

44. A method of modulating the activity of a polypeptide of any one of claims 20-30 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

45. A method for identifying a compound which modulates the activity of a polypeptide of any one of claims 20-30 comprising:

a) contacting a polypeptide of any one of claims 20-30 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

A

```
GCAGCATCCGACAGGAGCCCTGGCAAACAGGACGGATTTCCAGGACTCTACCAGCTGCCA    60
GACACGGCAGGGAGAGACCCCAGACCTCCTGGGTCCTGGCTGTGGGCCCGGATTGGGCTC   120
CCAAGTGGCGTTTGACTCACGTGGGGACACTCTTGGAAGAGACGACACCAGGAGCCTGAA   180
TGGGGAACGATTCTGTCAGCTACGAGTATGGGGATTACAGCGACCTCTCGGACCGCCCTG   240
TGGACTGCCTGGATGGCGCCTGCCTGGCCATCGACCCGCTGCGCGTGGCCCCGCTCCCAC   300
TGTATGCCGCCATCTTCCTGGTGGGGGTGCCGGGCAATGCCATGGTGGCCTGGGTGGCTG   360
GGAAGGTGGCCCGCCGGAGGGTGGGTGCCACCTGGTTGCTCCACCTGGCCGTGGCGGATT   420
TGCTGTGCTGTTTGTCTCTGCCCATCCTGGCAGTGCCCATTGCCCGTGGAGGCCACTGGC   480
CGTATGGTGCAGTGGGCTGTCGGGCGCTGCCCTCCATCATCCTGCTGACCATGTATGCCA   540
GCGTCCTGCTCCTGGCAGCTCTCAGTGCCGACCTCTGCTTCCTGGCTCTCGGGCCTGCCT   600
GGTGGTCTACGGTTCAGCGGGCGTGCGGGGTGCAGGTGGCCTGTGGGGCAGCCTGGACAC   660
TGGCCTTGCTGCTCACCGTGCCCTCCGCCATCTACCGCCGGCTGCACCAGGAGCACTTCC   720
CAGCCCGGCTGCAGTGTGTGGTGGACTACGGCGGCTCCTCCAGCACCGAGAATGCGGTGA   780
CTGCCATCCGGTTTCTTTTTGGCTTCCTGGGGCCCCTGGTGGCCGTGGCCAGCTGCCACA   840
GTGCCCTCCTGTGCTGGGCAGCCCGACGCTGCCGGCCGCTGGGCACAGCCATTGTGGTGG   900
GGTTTTTTGTCTGCTGGGCACCCTACCACCTGCTGGGGCTGGTGCTCACTGTGGCGGCCC   960
CGAACTCCGCACTCCTGGCCAGGGCCCTGCGGGCTGAACCCCTCATCGTGGGCCTTGCCC  1020
TCGCTCACAGCTGCCTCAATCCCATGCTCTTCCTGTATTTTGGGAGGGCTCAACTCCGCC  1080
GGTCACTGCCAGCTGCCTGTCACTGGGCCCTGAGGGAGTCCCAGGGCCAGGACGAAAGTG  1140
TGGACAGCAAGAAATCCACCAGCCATGACCTGGTCTCGGAGATGGAGGTGTAGGCTGGAG  1200
AGACATTGTGGGTGTGTATCTTCTTATCTCATTTCACAAGACTGGCTTCAGGCATAGCTG  1260
GATCCAGGAGCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1320
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA           1374
```

B

```
MGNDSVSYEYGDYSDLSDRPVDCLDGACLAIDPLRVAPLPLYAAIFLVGV    50
PGNAMVAWVAGKVARRRVGATWLLHLAVADLLCCLSLPILAVPIARGGHW   100
PYGAVGCRALPSIILLTMYASVLLLAALSADLCFLALGPAWWSTVQRACG   150
VQVACGAAWTLALLLTVPSAIYRRLHQEHFPARLQCVVDYGGSSSTENAV   200
TAIRFLFGFLGPLVAVASCHSALLCWAARRCRPLGTAIVVGFFVCWAPYH   250
LLGLVLTVAAPNSALLARALRAEPLIVGLALAHSCLNPMLFLYFGRAQLR   300
RSLPAACHWALRESQGQDESVDSKKSTSHDLVSEMEV                337
```

FIG. 1

Alignment of:

Sequence	Start	End	Sequencetype
Guinea Pig	(1 >	345)	PROTEIN
Human C5R1	(1 >	350)	PROTEIN
Human C5RL	(1 >	337)	PROTEIN

```
                1                                                   I    * 60
Guinea Pig      MM----VTVSYD-YDYNSTFLPDGFVDNY--VERLSFGDLVAVVIMVVVFLVGVPGNALV
Human C5R1      MNSFNYTTPDYGHYDDKDTLDLNTPVDKT--SNTLRVPDILALVIFAVVFLVGVLGNALV
Human C5RL      MGN-DSVSYEYGDYSDLSDRPVDCLDGACLAIDPLRVA---PLPLYAAIFLVGVPGNAMV

               61                  *   *  II                       *      120
Guinea Pig      VWVTACEARRHINAIWFLNLAAADLLSCLALPILLVSTVHLNHWYFGDTACKVLPSLILL
Human C5R1      VWVTAFEAKRTINAIWFLNLAVADFLSCLALPILFTSIVQHHHWPFGGAACSILPSLILL
Human C5RL      AWVAGKVARRRVGATWLLHLAVADLLCCLSLPILAVPIARGGHWPYGAVGCRALPSIILL

               121  III                                 *   IV   *     180
Guinea Pig      NMYTSILLLATISADRLLLVLSPIWCQRFRGGCLAWTACGLAWVLALLLSSPSFLYRRTH
Human C5R1      NMYASILLLATISADRFLLVFKPIWCQNFRGAGLAWIACAVAWGLALLLTIPSFLYRVVR
Human C5RL      TMYASVLLLAALSADLCFLALGPAWWSTVQRACGVQVACGAAWTLALLLTVPSAIYRRLH

               181      *                          *   V                240
Guinea Pig      NEHFSFKVYCVTDYGRDISKERAVALVRLLVGFIVPLITLTACYTFLLLRTWSRKATRSA
Human C5R1      EEYFPPKVLCGVDYSHDKRRERAVAIVRLVLGFLWPLLTLTICYTFILLRTWSRRATRST
Human C5RL      QEHFPARLQCVVDYGGSSSTENAVTAIRFLFGFLGPLVAVASCHSALL--CWAARRCRP-

               241         *  *  *  VI                                VII 300
Guinea Pig      KTVKVVVAVVSSFFVFWLPYQVTGILLAWHSPNSATYRNTKALDAVCVAFAYINCCINPI
Human C5R1      KTLKVVVAVVASFFIFWLPYQVTGIMMSFLEPSSPTFLLLNKLDSLCVSFAYINCCINPI
Human C5RL      ----LGTAIVVGFFVCWAPYHLLGLVLTVAAPNSALLARALRAEPLIVGLALAHSCLNPM

               301                                                 352
Guinea Pig      IYVVAGHGFQGRLLKSLPSVLRNVLTEESLDKRHQSFARSTVDTMPQKSESV
Human C5R1      IYVVAGQGFQGRLRKSLPSLLRNVLTEESVVRESKSFTRSTVDTMAQKTQAV
Human C5RL      LFLYFGRA---QLRRSLPAACHWALRESQGQDESVDSKKSTSHDLVSEME-V
```

FIG. 2

SEQUENCE LISTING

<110> Genetics Institute Inc. et al.

<120> NOVEL G PROTEIN-COUPLED RECEPTOR PROTEIN AND NUCLEIC ACID MOLECULES AND USES THEREFOR

<130> GNN-023PC

<140>
<141>

<150> 60/246,553
<151> 2000-11-07

<160> 5

<170> PatentIn Ver. 2.0

<210> 1
<211> 1374
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (180)..(1190)

<400> 1
gcagcatccg acaggagccc tggcaaacag gacggatttc caggactcta ccagctgcca 60

gacacggcag ggagagaccc cagacctcct gggtcctggc tgtgggcccg gattgggctc 120

ccaagtggcg tttgactcac gtggggacac tcttggaaga gacgacacca ggagcctga 179

atg ggg aac gat tct gtc agc tac gag tat ggg gat tac agc gac ctc 227
Met Gly Asn Asp Ser Val Ser Tyr Glu Tyr Gly Asp Tyr Ser Asp Leu
1 5 10 15

tcg gac cgc cct gtg gac tgc ctg gat ggc gcc tgc ctg gcc atc gac 275
Ser Asp Arg Pro Val Asp Cys Leu Asp Gly Ala Cys Leu Ala Ile Asp
20 25 30

ccg ctg cgc gtg gcc ccg ctc cca ctg tat gcc gcc atc ttc ctg gtg 323
Pro Leu Arg Val Ala Pro Leu Pro Leu Tyr Ala Ala Ile Phe Leu Val
35 40 45

ggg gtg ccg ggc aat gcc atg gtg gcc tgg gtg gct ggg aag gtg gcc 371
Gly Val Pro Gly Asn Ala Met Val Ala Trp Val Ala Gly Lys Val Ala
50 55 60

cgc cgg agg gtg ggt gcc acc tgg ttg ctc cac ctg gcc gtg gcg gat 419
Arg Arg Arg Val Gly Ala Thr Trp Leu Leu His Leu Ala Val Ala Asp
65 70 75 80

ttg ctg tgc tgt ttg tct ctg ccc atc ctg gca gtg ccc att gcc cgt 467
Leu Leu Cys Cys Leu Ser Leu Pro Ile Leu Ala Val Pro Ile Ala Arg
85 90 95

```
gga ggc cac tgg ccg tat ggt gca gtg ggc tgt cgg gcg ctg ccc tcc   515
Gly Gly His Trp Pro Tyr Gly Ala Val Gly Cys Arg Ala Leu Pro Ser
            100                 105                 110

atc atc ctg ctg acc atg tat gcc agc gtc ctg ctc ctg gca gct ctc   563
Ile Ile Leu Leu Thr Met Tyr Ala Ser Val Leu Leu Leu Ala Ala Leu
        115                 120                 125

agt gcc gac ctc tgc ttc ctg gct ctc ggg cct gcc tgg tgg tct acg   611
Ser Ala Asp Leu Cys Phe Leu Ala Leu Gly Pro Ala Trp Trp Ser Thr
    130                 135                 140

gtt cag cgg gcg tgc ggg gtg cag gtg gcc tgt ggg gca gcc tgg aca   659
Val Gln Arg Ala Cys Gly Val Gln Val Ala Cys Gly Ala Ala Trp Thr
145                 150                 155                 160

ctg gcc ttg ctg ctc acc gtg ccc tcc gcc atc tac cgc cgg ctg cac   707
Leu Ala Leu Leu Leu Thr Val Pro Ser Ala Ile Tyr Arg Arg Leu His
                165                 170                 175

cag gag cac ttc cca gcc cgg ctg cag tgt gtg gtg gac tac ggc ggc   755
Gln Glu His Phe Pro Ala Arg Leu Gln Cys Val Val Asp Tyr Gly Gly
            180                 185                 190

tcc tcc agc acc gag aat gcg gtg act gcc atc cgg ttt ctt ttt ggc   803
Ser Ser Ser Thr Glu Asn Ala Val Thr Ala Ile Arg Phe Leu Phe Gly
        195                 200                 205

ttc ctg ggg ccc ctg gtg gcc gtg gcc agc tgc cac agt gcc ctc ctg   851
Phe Leu Gly Pro Leu Val Ala Val Ala Ser Cys His Ser Ala Leu Leu
    210                 215                 220

tgc tgg gca gcc cga cgc tgc cgg ccg ctg ggc aca gcc att gtg gtg   899
Cys Trp Ala Ala Arg Arg Cys Arg Pro Leu Gly Thr Ala Ile Val Val
225                 230                 235                 240

ggg ttt ttt gtc tgc tgg gca ccc tac cac ctg ctg ggg ctg gtg ctc   947
Gly Phe Phe Val Cys Trp Ala Pro Tyr His Leu Leu Gly Leu Val Leu
                245                 250                 255

act gtg gcg gcc ccg aac tcc gca ctc ctg gcc agg gcc ctg cgg gct   995
Thr Val Ala Ala Pro Asn Ser Ala Leu Leu Ala Arg Ala Leu Arg Ala
            260                 265                 270

gaa ccc ctc atc gtg ggc ctt gcc ctc gct cac agc tgc ctc aat ccc   1043
Glu Pro Leu Ile Val Gly Leu Ala Leu Ala His Ser Cys Leu Asn Pro
        275                 280                 285

atg ctc ttc ctg tat ttt ggg agg gct caa ctc cgc cgg tca ctg cca   1091
Met Leu Phe Leu Tyr Phe Gly Arg Ala Gln Leu Arg Arg Ser Leu Pro
    290                 295                 300

gct gcc tgt cac tgg gcc ctg agg gag tcc cag ggc cag gac gaa agt   1139
Ala Ala Cys His Trp Ala Leu Arg Glu Ser Gln Gly Gln Asp Glu Ser
305                 310                 315                 320

gtg gac agc aag aaa tcc acc agc cat gac ctg gtc tcg gag atg gag   1187
Val Asp Ser Lys Lys Ser Thr Ser His Asp Leu Val Ser Glu Met Glu
                325                 330                 335
```

gtg taggctggag agacattgtg ggtgtgtatc ttcttatctc atttcacaag 1240
Val

actggcttca ggcatagctg gatccaggag ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa 1300

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1360

aaaaaaaaaa aaaa 1374

<210> 2
<211> 337
<212> PRT
<213> Homo sapiens

<400> 2

Met Gly Asn Asp Ser Val Ser Tyr Glu Tyr Gly Asp Tyr Ser Asp Leu
1 5 10 15

Ser Asp Arg Pro Val Asp Cys Leu Asp Gly Ala Cys Leu Ala Ile Asp
20 25 30

Pro Leu Arg Val Ala Pro Leu Pro Leu Tyr Ala Ala Ile Phe Leu Val
35 40 45

Gly Val Pro Gly Asn Ala Met Val Ala Trp Val Ala Gly Lys Val Ala
50 55 60

Arg Arg Arg Val Gly Ala Thr Trp Leu Leu His Leu Ala Val Ala Asp
65 70 75 80

Leu Leu Cys Cys Leu Ser Leu Pro Ile Leu Ala Val Pro Ile Ala Arg
85 90 95

Gly Gly His Trp Pro Tyr Gly Ala Val Gly Cys Arg Ala Leu Pro Ser
100 105 110

Ile Ile Leu Leu Thr Met Tyr Ala Ser Val Leu Leu Leu Ala Ala Leu
115 120 125

Ser Ala Asp Leu Cys Phe Leu Ala Leu Gly Pro Ala Trp Trp Ser Thr
130 135 140

Val Gln Arg Ala Cys Gly Val Gln Val Ala Cys Gly Ala Ala Trp Thr
145 150 155 160

Leu Ala Leu Leu Leu Thr Val Pro Ser Ala Ile Tyr Arg Arg Leu His
165 170 175

Gln Glu His Phe Pro Ala Arg Leu Gln Cys Val Val Asp Tyr Gly Gly
180 185 190

Ser Ser Ser Thr Glu Asn Ala Val Thr Ala Ile Arg Phe Leu Phe Gly
195 200 205

Phe Leu Gly Pro Leu Val Ala Val Ala Ser Cys His Ser Ala Leu Leu
210 215 220

Cys Trp Ala Ala Arg Arg Cys Arg Pro Leu Gly Thr Ala Ile Val Val
225 230 235 240

```
Gly Phe Phe Val Cys Trp Ala Pro Tyr His Leu Leu Gly Leu Val Leu
                245                 250                 255

Thr Val Ala Ala Pro Asn Ser Ala Leu Leu Ala Arg Ala Leu Arg Ala
            260                 265                 270

Glu Pro Leu Ile Val Gly Leu Ala Leu Ala His Ser Cys Leu Asn Pro
        275                 280                 285

Met Leu Phe Leu Tyr Phe Gly Arg Ala Gln Leu Arg Arg Ser Leu Pro
    290                 295                 300

Ala Ala Cys His Trp Ala Leu Arg Glu Ser Gln Gly Gln Asp Glu Ser
305                 310                 315                 320

Val Asp Ser Lys Lys Ser Thr Ser His Asp Leu Val Ser Glu Met Glu
                325                 330                 335

Val
```

<210> 3
<211> 1011
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1011)

<400> 3

```
atg ggg aac gat tct gtc agc tac gag tat ggg gat tac agc gac ctc   48
Met Gly Asn Asp Ser Val Ser Tyr Glu Tyr Gly Asp Tyr Ser Asp Leu
  1               5                  10                  15

tcg gac cgc cct gtg gac tgc ctg gat ggc gcc tgc ctg gcc atc gac   96
Ser Asp Arg Pro Val Asp Cys Leu Asp Gly Ala Cys Leu Ala Ile Asp
            20                  25                  30

ccg ctg cgc gtg gcc ccg ctc cca ctg tat gcc gcc atc ttc ctg gtg   144
Pro Leu Arg Val Ala Pro Leu Pro Leu Tyr Ala Ala Ile Phe Leu Val
        35                  40                  45

ggg gtg ccg ggc aat gcc atg gtg gcc tgg gtg gct ggg aag gtg gcc   192
Gly Val Pro Gly Asn Ala Met Val Ala Trp Val Ala Gly Lys Val Ala
    50                  55                  60

cgc cgg agg gtg ggt gcc acc tgg ttg ctc cac ctg gcc gtg gcg gat   240
Arg Arg Arg Val Gly Ala Thr Trp Leu Leu His Leu Ala Val Ala Asp
65                  70                  75                  80

ttg ctg tgc tgt ttg tct ctg ccc atc ctg gca gtg ccc att gcc cgt   288
Leu Leu Cys Cys Leu Ser Leu Pro Ile Leu Ala Val Pro Ile Ala Arg
                85                  90                  95

gga ggc cac tgg ccg tat ggt gca gtg ggc tgt cgg gcg ctg ccc tcc   336
Gly Gly His Trp Pro Tyr Gly Ala Val Gly Cys Arg Ala Leu Pro Ser
            100                 105                 110
```

```
atc atc ctg ctg acc atg tat gcc agc gtc ctg ctc ctg gca gct ctc   384
Ile Ile Leu Leu Thr Met Tyr Ala Ser Val Leu Leu Leu Ala Ala Leu
        115                 120                 125

agt gcc gac ctc tgc ttc ctg gct ctc ggg cct gcc tgg tgg tct acg   432
Ser Ala Asp Leu Cys Phe Leu Ala Leu Gly Pro Ala Trp Trp Ser Thr
    130                 135                 140

gtt cag cgg gcg tgc ggg gtg cag gtg gcc tgt ggg gca gcc tgg aca   480
Val Gln Arg Ala Cys Gly Val Gln Val Ala Cys Gly Ala Ala Trp Thr
145                 150                 155                 160

ctg gcc ttg ctg ctc acc gtg ccc tcc gcc atc tac cgc cgg ctg cac   528
Leu Ala Leu Leu Leu Thr Val Pro Ser Ala Ile Tyr Arg Arg Leu His
                165                 170                 175

cag gag cac ttc cca gcc cgg ctg cag tgt gtg gtg gac tac ggc ggc   576
Gln Glu His Phe Pro Ala Arg Leu Gln Cys Val Val Asp Tyr Gly Gly
            180                 185                 190

tcc tcc agc acc gag aat gcg gtg act gcc atc cgg ttt ctt ttt ggc   624
Ser Ser Ser Thr Glu Asn Ala Val Thr Ala Ile Arg Phe Leu Phe Gly
        195                 200                 205

ttc ctg ggg ccc ctg gtg gcc gtg gcc agc tgc cac agt gcc ctc ctg   672
Phe Leu Gly Pro Leu Val Ala Val Ala Ser Cys His Ser Ala Leu Leu
    210                 215                 220

tgc tgg gca gcc cga cgc tgc cgg ccg ctg ggc aca gcc att gtg gtg   720
Cys Trp Ala Ala Arg Arg Cys Arg Pro Leu Gly Thr Ala Ile Val Val
225                 230                 235                 240

ggg ttt ttt gtc tgc tgg gca ccc tac cac ctg ctg ggg ctg gtg ctc   768
Gly Phe Phe Val Cys Trp Ala Pro Tyr His Leu Leu Gly Leu Val Leu
                245                 250                 255

act gtg gcg gcc ccg aac tcc gca ctc ctg gcc agg gcc ctg cgg gct   816
Thr Val Ala Ala Pro Asn Ser Ala Leu Leu Ala Arg Ala Leu Arg Ala
            260                 265                 270

gaa ccc ctc atc gtg ggc ctt gcc ctc gct cac agc tgc ctc aat ccc   864
Glu Pro Leu Ile Val Gly Leu Ala Leu Ala His Ser Cys Leu Asn Pro
        275                 280                 285

atg ctc ttc ctg tat ttt ggg agg gct caa ctc cgc cgg tca ctg cca   912
Met Leu Phe Leu Tyr Phe Gly Arg Ala Gln Leu Arg Arg Ser Leu Pro
    290                 295                 300

gct gcc tgt cac tgg gcc ctg agg gag tcc cag ggc cag gac gaa agt   960
Ala Ala Cys His Trp Ala Leu Arg Glu Ser Gln Gly Gln Asp Glu Ser
305                 310                 315                 320

gtg gac agc aag aaa tcc acc agc cat gac ctg gtc tcg gag atg gag   1008
Val Asp Ser Lys Lys Ser Thr Ser His Asp Leu Val Ser Glu Met Glu
                325                 330                 335

gtg                                                               1011
Val
```

<210> 4
<211> 345
<212> PRT
<213> Cavia porcellus

<400> 4

```
Met Met Val Thr Val Ser Tyr Asp Tyr Asp Tyr Asn Ser Thr Phe Leu
  1               5                  10                  15

Pro Asp Gly Phe Val Asp Asn Tyr Val Glu Arg Leu Ser Phe Gly Asp
             20                  25                  30

Leu Val Ala Val Val Ile Met Val Val Val Phe Leu Val Gly Val Pro
         35                  40                  45

Gly Asn Ala Leu Val Val Trp Val Thr Ala Cys Glu Ala Arg Arg His
     50                  55                  60

Ile Asn Ala Ile Trp Phe Leu Asn Leu Ala Ala Ala Asp Leu Leu Ser
 65                  70                  75                  80

Cys Leu Ala Leu Pro Ile Leu Leu Val Ser Thr Val His Leu Asn His
                 85                  90                  95

Trp Tyr Phe Gly Asp Thr Ala Cys Lys Val Leu Pro Ser Leu Ile Leu
            100                 105                 110

Leu Asn Met Tyr Thr Ser Ile Leu Leu Leu Ala Thr Ile Ser Ala Asp
        115                 120                 125

Arg Leu Leu Leu Val Leu Ser Pro Ile Trp Cys Gln Arg Phe Arg Gly
    130                 135                 140

Gly Cys Leu Ala Trp Thr Ala Cys Gly Leu Ala Trp Val Leu Ala Leu
145                 150                 155                 160

Leu Leu Ser Ser Pro Ser Phe Leu Tyr Arg Arg Thr His Asn Glu His
                165                 170                 175

Phe Ser Phe Lys Val Tyr Cys Val Thr Asp Tyr Gly Arg Asp Ile Ser
            180                 185                 190

Lys Glu Arg Ala Val Ala Leu Val Arg Leu Leu Val Gly Phe Ile Val
        195                 200                 205

Pro Leu Ile Thr Leu Thr Ala Cys Tyr Thr Phe Leu Leu Leu Arg Thr
    210                 215                 220

Trp Ser Arg Lys Ala Thr Arg Ser Ala Lys Thr Val Lys Val Val Val
225                 230                 235                 240

Ala Val Val Ser Ser Phe Phe Val Phe Trp Leu Pro Tyr Gln Val Thr
                245                 250                 255

Gly Ile Leu Leu Ala Trp His Ser Pro Asn Ser Ala Thr Tyr Arg Asn
            260                 265                 270

Thr Lys Ala Leu Asp Ala Val Cys Val Ala Phe Ala Tyr Ile Asn Cys
        275                 280                 285
```

```
Cys Ile Asn Pro Ile Ile Tyr Val Val Ala Gly His Gly Phe Gln Gly
    290                 295                 300

Arg Leu Leu Lys Ser Leu Pro Ser Val Leu Arg Asn Val Leu Thr Glu
305                 310                 315                 320

Glu Ser Leu Asp Lys Arg His Gln Ser Phe Ala Arg Ser Thr Val Asp
                325                 330                 335

Thr Met Pro Gln Lys Ser Glu Ser Val
            340                 345
```

<210> 5
<211> 350
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Asn Ser Phe Asn Tyr Thr Thr Pro Asp Tyr Gly His Tyr Asp Asp
  1               5                  10                  15

Lys Asp Thr Leu Asp Leu Asn Thr Pro Val Asp Lys Thr Ser Asn Thr
             20                  25                  30

Leu Arg Val Pro Asp Ile Leu Ala Leu Val Ile Phe Ala Val Val Phe
         35                  40                  45

Leu Val Gly Val Leu Gly Asn Ala Leu Val Val Trp Val Thr Ala Phe
     50                  55                  60

Glu Ala Lys Arg Thr Ile Asn Ala Ile Trp Phe Leu Asn Leu Ala Val
 65                  70                  75                  80

Ala Asp Phe Leu Ser Cys Leu Ala Leu Pro Ile Leu Phe Thr Ser Ile
                85                  90                  95

Val Gln His His His Trp Pro Phe Gly Gly Ala Ala Cys Ser Ile Leu
            100                 105                 110

Pro Ser Leu Ile Leu Leu Asn Met Tyr Ala Ser Ile Leu Leu Leu Ala
        115                 120                 125

Thr Ile Ser Ala Asp Arg Phe Leu Leu Val Phe Lys Pro Ile Trp Cys
    130                 135                 140

Gln Asn Phe Arg Gly Ala Gly Leu Ala Trp Ile Ala Cys Ala Val Ala
145                 150                 155                 160

Trp Gly Leu Ala Leu Leu Leu Thr Ile Pro Ser Phe Leu Tyr Arg Val
                165                 170                 175

Val Arg Glu Glu Tyr Phe Pro Pro Lys Val Leu Cys Gly Val Asp Tyr
            180                 185                 190

Ser His Asp Lys Arg Arg Glu Arg Ala Val Ala Ile Val Arg Leu Val
        195                 200                 205

Leu Gly Phe Leu Trp Pro Leu Leu Thr Leu Thr Ile Cys Tyr Thr Phe
    210                 215                 220
```

```
Ile Leu Leu Arg Thr Trp Ser Arg Arg Ala Thr Arg Ser Thr Lys Thr
225                 230                 235                 240

Leu Lys Val Val Val Ala Val Val Ala Ser Phe Phe Ile Phe Trp Leu
                245                 250                 255

Pro Tyr Gln Val Thr Gly Ile Met Met Ser Phe Leu Glu Pro Ser Ser
            260                 265                 270

Pro Thr Phe Leu Leu Leu Asn Lys Leu Asp Ser Leu Cys Val Ser Phe
        275                 280                 285

Ala Tyr Ile Asn Cys Cys Ile Asn Pro Ile Ile Tyr Val Val Ala Gly
    290                 295                 300

Gln Gly Phe Gln Gly Arg Leu Arg Lys Ser Leu Pro Ser Leu Leu Arg
305                 310                 315                 320

Asn Val Leu Thr Glu Glu Ser Val Val Arg Glu Ser Lys Ser Phe Thr
                325                 330                 335

Arg Ser Thr Val Asp Thr Met Ala Gln Lys Thr Gln Ala Val
            340                 345                 350
```